Machine Learning Methods to Predict Wind Intensity

António Silva, Fernando Aristides
Departamento de Engenharia Mecânica
Instituto Superior de Engenharia do Porto
Rua Dr. António Bernardino de Almeida, 431
4200-072 Porto, Portugal
{afs, fac}@isep.ipp.pt

Rui Camacho, Nuno Fidalgo
INESC & FEUP
Universidade do Porto
Rua Dr Roberto Frias, s/n
4200-465 Porto, Portugal
rcamacho@fe.up.pt, jfidalgo@inescporto.pt

Abstract: A good prediction of wind intensity is crucial for the optimal management of wind-based energy production. It is important to be able to predict how much energy we can expect from wind mills in order to manage a country wide electrical production system. In this study we compare the performance of two approaches. The first approach is a hybrid system using Neural Networks, improved with Genetic Algorithms to tune the subset of variables and the Neural Network parameters. The second approach uses a Model Tree constructor called M5. The comparative study uses a data set of real wind intensity measurements. The results show that both methods produced highly accurate prediction with the ANN having better performance.

Key–Words: Wind velocity prediction, Neural Networks, Model Trees

1 Introduction

It is nowadays agreed that oil reserves will be extinguished very soon. A new paradigm for sustainable energy source is therefore required. It is also believed that "renewable energies" are a serious alternative to replace oil and other fossil energy sources like coal. "renewable energies" have very appealing challenges namely they produce much less pollution (emissions of CO2 for example) than the traditional ones.

After the 70s, of last century, there has been an increased interest in the use of renewable energies. Among the different kinds of renewable energies, wind energy is regarded as one of the most promising. Wind technology is nowadays a mature technology specially in Europe and USA. Wind energy offers a viable and economical alternative to conventional power plants in many areas of the country. Wind is a clean fuel; wind farms produce no air or water pollution because no fuel is burned. The most serious environmental drawbacks to wind machines may be their negative effect on wild bird populations and the visual impact on the landscape.

It is a priority in almost every country to develop alternative energy production sources. Eolic energy is getting an increased attention and developed in recent years. According to [2] it is expected that between 2500 and 3000 MW of wind mills will be installed in Portugal alone until 2010.

Despite the increased use of wind energy there is almost no tools to help in the planning of new infrastructures and in the systematic identification of places with high potential for wind energy production. It would be very useful to have, for each country, a map with the potential of wind energy production for each region. To achieve such goal one needs to have deep knowledge of wind characteristics in each region and specially its intensity. This knowledge is crucial not only to determine the energy potential of the wind in each location but also to be able to plan future electric networks and related infrastructures.

Evaluating the potential of a region for eolic energy production is done using different methodologies. The European Wind Atlas/Wasp [9] uses the so called classical methodology. This methodology uses the following parameters: average of wind velocity, direction, Weibull distribution, daily wind profile, power fluxes and; annual energy estimation. Other equally used methodologies (such as the spectral method) also use a large amount of other information essential to select and characterise the eolic potential and energy production of the wind.

We propose the use of computational-based wind intensity predictors that are constructed using historical records of wind intensity collected at each region. Our approach is based on a hybrid model that combines Artificial Neural Networks (ANN) and a Genetic Algorithm (GA). The GA is used to choose the most promising subset of variables to make wind velocity prediction. The AN is used as the fitness function (evaluating function) of the GA. The model is evaluated in real world data from the region of Évora in Portugal and is compared with a Model Tree tool called M5. M5 is known to be a good regression algo-

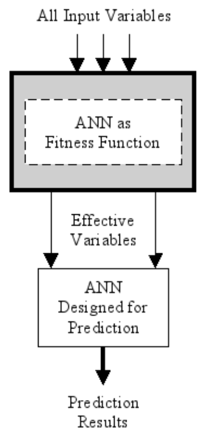

Figure 1: Hybrid model combining Genetic Algorithms and Artificial Neural Networks.

rithm.

The rest of the paper is organised as follows. Section 2 describes our hybrid model proposal. Section 3 presents the Model Tree tool with which we have compared our model. Section 4 introduces the reader to the pre-processing of the raw data that includes the selection of relevant variables for wind velocity prediction. In Section 5 we describe the experiments we performed and discuss the results obtained. The conclusions are drawn in Section 6.

2 The hybrid model

The model proposed in this paper is schematically shown in Figure 1. It includes a Genetic Algorithm and an Artificial Neural Network. Some authors like [1] argue that an hybrid model increases the predictive power of of the tool. We therefore make an assessment study where we compare the ANN with and without the GA and also compare the hybrid model with Model Trees that usually give good results in regression problems.

In the hybrid model, all input variables are provided to the GA that performs a kind of heuristic search to optimise the combination of variables producing the best predictive error. In short, the GA is used to solve a feature subset selection problem. During the heuristic search an ANN is used as the fitness function of the GA, that is, for each chromosome an ANN is trained and used to evaluate the combination of variables encoded in the chromosome. Once the best subset of variables is determined a ANN is then used to make the wind velocity predictions.

2.1 The Genetic Algorithm

In our data set for wind intensity prediction the number of given variables is quite large. We have to solve a feature (subset) selection problem in order to determine the most influential ones. The effect of this features selection procedure is three fold: identify the most relevant variables; speed-up the predictive process and; reduce the chance of over-fitting the data.

Genetic Algorithms are Machine Learning algorithms that use the metaphor of Evolution of Species and Natural Selection law in optimization problems. As GAs use special encoding schemes for the parameters they are designed to tune, we have adopted the following encoding. A chromosome is an array of binary digits representing the different variables of the data set. The number of genes in the chromosome is equal to the number of variables to analyse. The binary digit 1 in position **i** of a chromosome indicates that the variable number **i** is in the subset. On the other hand the binary digit 0 (zero) at position **i** of a chromosome indicates that variable number **i** is not in the subset represented by the chromosome. For example, the chromosome [1 0 0 1 0 1] represents the choice for the first, forth and sixth variables as effective variables. The GA algorithm starts with a population of individuals (represented by a set of chromosomes) each individual is a potential solution to the problem at hands. As in Nature each population may produce an off-spring by combination of genetic material of the individuals in the population. The Natural Selection law in a GA is implemented by discarding the individuals less fit in the population and also by allowing only the most fit to participate in the generation of the off-spring. In our implementation we have used the GA single-point crossover and the single-point mutation operations. In a GA the process of producing off-springs, selecting the most fit is an iterative process, maintaining the size of the population fixed and terminating only after a fixed set of cycles or when the fitness function reaches a desirable value. Each individual is evaluated using a fitness function. In our implementation the fitness function is an Artificial Neural Network trained with the input variables determined by the chromosome encoding.

2.2 The Artificial Neural Network

According to Preschelt ([7]) connectionist models that are able to learn using historical data are adequate tools for predicting future situation where noise and incomplete information are present. In our work we used an Artificial Neural Network of the type Multilayer Perceptron (MLP) as the fitness function for the Genetic Algorithm. The training of the ANN was done using the Levenberg-Marquard algorithm. We

MSE	**Max. Error**	Error Sum
$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$max_{i=1\ \dots\ n}\lvert y_i - \hat{y}_i\rvert$	$\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$

Table 1: Error measures used to assess the effectiveness of the hybrid model.

have used the MATLAB ([3]) Toolbox for ANN. We have conducted preliminary tests to determine the best number of neurons in the hidden layer and obtained the number of 6 neurons for the hidden layer. This number of neurons was used throughout all experiments. The number of neurons on the input layer varied according to the problem at hands. The output layer has only one neuron. The error measures used in the experiments are defined in Table 1.

where y_i represents the actual value of the output variable, $\hat{y}_i$ the value predicted by the model and n is the number of registers.

3 Model Trees

We have compared the hybrid model with a Model Tree constructor called M5 ([8]) as implemented in Weka's M5P algorithm ([10]). A Model Tree is like a Decision Tree with different kind of leaves. As in a Decision Tree the root and the internal nodes encode a test on an attribute. For each possible outcome of the test there is a branch to a child node. The main difference between a Model Tree and a Decision Tree concern the leaves. Whereas in a Decision Tree a leaf stores the class value to assign to a object that reaches that leaf, in a Model Trees there is a linear equation on the attributes that assigns a numerical value to the object that reaches that leaf. To predict the value for a new object one start at the root and follows a path to a leaf. At each node of the path the corresponding test is performed and the outcome determines the child that follows in the path. When reaching a leaf the equation stored there is computed and the result is the value predicted for that object.

4 Data Pre-processing

When predicting wind intensity there are usually two possible situations we have to address: absence or near absence of meteorological information of the region or; overwhelming amount of information. According to Kalogirou ([4]) we should address the first situation by using meteorological information from neiborhood regions. The second situation we have to still address the two questions: is all of the information relevant? or is there a minimal subset of the information containing all relevant information?

In our study we used data from the meteorological observatory of Mitra (vora – Portugal). Data collected at this observatory include: air temperature; relative air humidity; wind intensity and direction; rain quantity and; solar radiation. All these measurements are done each 10 minutes and is done since 2002 up today. To this information we may add average, maximum and minimum values each hour/day/month of both wind intensity and direction.

Our goal in this study is to make predictions for the next 24 hours. The traditional way of doing such prediction is to use statistical methods or to look at historical data. Recently ANN have been used successfully in prediction problems ([5]). However, using a large number of input variable to a ANN increases substantially the training time, the error rate may also increase due to the use of inappropriate variables and time is spent to collect data than may not be relevant to the prediction task.

More recently Makvandi et al. ([6]) propose that only relevant variable be used as input variables to train an ANN. We propose the use of a GA as an heuristic approach to select the subset of most relevant input variables.

5 The experiments

The data we have used in our study is from the meteorological observatory of Mitra (vora – Portugal). Data collected at this observatory include: air temperature; relative air humidity; wind intensity and direction; rain quantity and; solar radiation. All these measurements are done each 10 minutes and is done since 2002 up today. The tools used in the experiments are: the ANN as implemented in Matlab Toolbox and; Weka version 3.4.13 implementation of Model Tree learner M5. M5 was used with the default parameter values.

The goal of the experiments is to make prediction for 24 hours (1 full day). We setup two sets of experiments. First we define a simpler problem with 10 input variables and then we used 24 input variables. For the two sets of experiments we assessed how the GA performed in the identification of the most relevant input variables. We run the models with all the input variables and then with the GA as the feature selection prep-processing. We have used the 1st of April 2007 as the day for which we made the wind intensity predictions. This day of our choice has a considerable variation in wind intensity and is therefore a major challenge for the models under evaluation.

1. Wind intensity
2. Wind direction
3. Air temperature
4. Air relative humidity
5. Intensity change in the previous hour
6. Direction change in the previous hour
7. Temperature change in the previous hour
8. Humidity change in the previous hour
9. Difference between the wind intensity value and the day average
10. Strong change in wind intensity (+ or 3m/s)

Table 2: Initial attribute set used in the experiments.

	MSE	Max. Error	Error Sum
without GA	2.32	3.01	32.34
with GA	1.18	1.78	20.63

Table 3: Results obtained without using GA and when using the GA for feature selection.

5.1 24 hour prediction

To predict the wind intensity each hour in 1st April 2007 we used the 10 variables described in Table 2. To make this prediction we only use data from the previous 24 hours. To make the prediction we used the data of the day before the predicted day. The results of the experiments using 10 variables can be seen in Figure 2 and Table 3 for the case of not using the GA and in Figure 3 and Table 3 for the case where GA was used. In this experiment the GA selected 5 variables out of the initial 10. The results indicate that the use of a GA may improve the predictive power of the tool. In this experiment the most fit chromosome was: [0 0 0 1 1 0 1 0 1 1] meaning that only 5 variables were effective in the prediction of wind intensity.

We have carried out a more complex experiment were the tool was provided with 24 variable, 14 more than the previous 10. The extra variables were: minimum, maximum and average of day wind intensity; minimum, maximum and average of day wind direction; minimum, maximum and average day temperature; minimum, maximum and average day humidity; average from beginning of day up the previous hour of the prediction for wind intensity and ; average from beginning of day up the previous hour of the prediction for wind direction.

Using the 24 input variables the GA choose 11 and reduced the error as can be seen in Table 4. As previously said we have used a Model Tree (M5 algorithm) as a comparison tool. The results of using M5 with the 24 input variables is presented in Table 4. M5 constructed a model tree with 909 leaves.

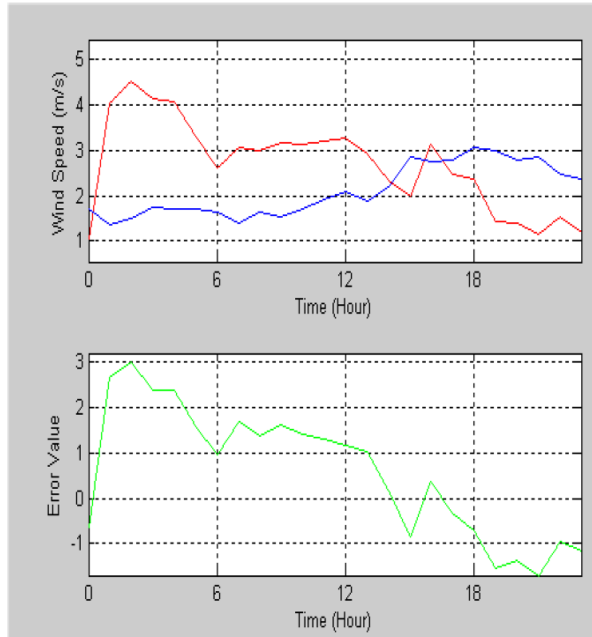

Figure 2: 24 hours prediction without using the Genetic Algorithm.

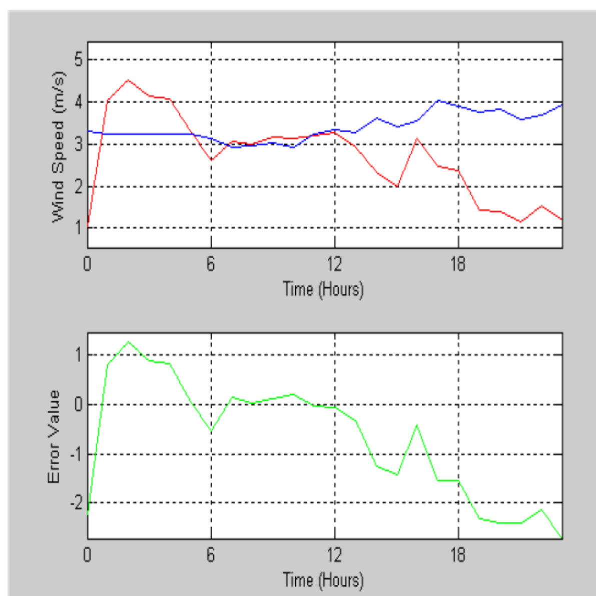

Figure 3: 24 hours prediction using the model combining Genetic Algorithms and Artificial Neural Networks.

tool	MSE	Max. Error	Error Sum
ANN & GA	1.18	1.78	20.63
M5	2.82	3.74	32.2

Table 4: Results obtained when using 24 input variables. First line is the result with ANN & GA. Second line is the result of using M5.

6 Conclusion

We have presented a hybrid model for wind intensity prediction. The hybrid model is based on a combination of Genetic Algorithms with Artificial Neural Networks. The goal of the GA is to select the most relevant variable for win intensity prediction.

We have compare our proposed model with a well known regression tree algorithm called M5.

The results show that the GA is effective in selecting relevant variables and achieves prediction values that are considered quite good by specialists. The proposed model also compares very well with the M5 results.

We think that the proposed model will be a valuable tool for tasks of planning new locations of eolic parks since those tasks need to be able to predict the eolic energy potential of each region.

Acknowledgements:

The authors would like to thank to “Centro de Geofísica de Évora, specially to Professor Ana Maria Silva for providing us with the data.

References:

[1] Alvaro V. *A hybrid linear-neural model for time series forecasting*, IEEE Transactions on Neural Network, V(11), pp.1402-1412, 2000.

[2] Castro R. *Energias Renováveis e Produção Descentralizada Introdução á Energia Eólica*, Universidade Tcnica de Lisboa Texto não publicado, Instituto Superior Técnico da Universidade Técnica de Lisboa, Lisboa, 2003.

[3] Demuth H. and Beale M. *Neural Network Toolbox for Use with MATLAB Users Guide*, Version 4, MathWoks, Inc. 2003.

[4] Kalogirou S., Neocleous C., Pashiardis S. and Schizas C, *Wind Speed Prediction Using Artificial Neural Networks*, Proceedings of the European Symposium on Intelligent Techniques. ESIT99, Crete, Greece, 1999.

[5] Kariniotakis G., Pinson P., and Siebert N, *The State of the Art in Short-term Prediction of Wind Power From an Offshore Perspective*, Proceedings of 2004 SeaTechWeek. Brest, France, 2004.

[6] Makvandi P., Jassbi J. and Khanmohammadi S., *Application of Genetic Algorithm and Neural Network in Forecasting with Good Data*, Proceedings of the 6th WSEAS Int. Conf. on Neural Networks , Lisboa, Portugal, June 16-18, pp.56-61, 2005.

[7] Preschelt L., *PROBEN1 A Set of Neural Network Benchmark Problems and Benchmarking Rules*, Research Report. Fakultt fr Informatik, Universitt Karlsruhe, Germany, 1994.

[8] Quinlan J. R. *Learning with continuous classes*, Proceedings of the Australian Joint Conference on Artificial Intelligence. 343–348. World Scientific, Singapore, 1992.

[9] WASP (2008). http://www.wasp.dk

[10] Ian H. Witten and Eibe Frank, *Data Mining: Practical machine learning tools and techniques*, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.